

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-142485

(43)Date of publication of application : 25.05.2001

(51)Int.Cl. G10L 15/24
G10L 11/02
G10L 15/04
G10L 15/10
G10L 15/22

(21)Application number : 11-326753

(71)Applicant : SHARP CORP

(22)Date of filing : 17.11.1999

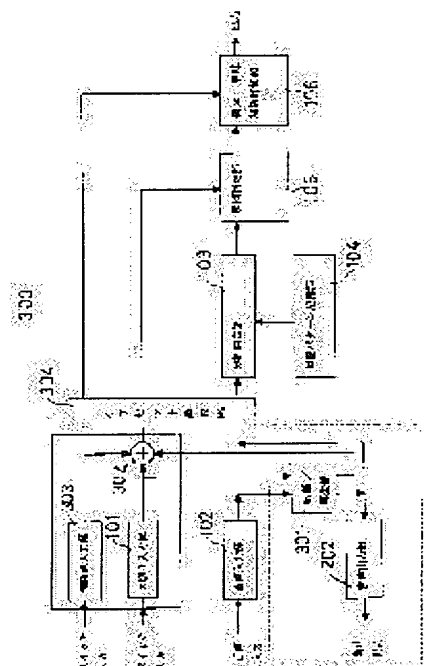
(72)Inventor : TODA MANABU

(54) METHOD AND DEVICE FOR VOICE RECOGNITION AND PORTABLE TELEPHONE SET

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce an erroneous discrimination to improve the recognition rate by properly adjusting the timing of punctuation input of words in the case of voice recognition.

SOLUTION: A voice input is converted into digital data by a voice input part 102 and is stored in a recording/reproducing part 301 together with a punctuation input inputted to a punctuation input part 101. Voice recognition is performed by an analysis and collation part 103, a candidate discrimination part 105, and a syntactic and semantic analysis and evaluation part 106 while reproducing recorded voice data. New punctuation inputting can be performed or punctuation marks can be inputted through a punctuation mark input part 303 at the time of reproducing. The punctuation input and the punctuation mark input are given to the candidate discrimination part 105 through a punctuation addition part 302. When a switch input is operated, delay due to a debounce processing and the time lag between voice input and switch operation are compensated. Evaluation of voice recognition candidates is changed by temporal compensation at the input time, and this change is reflected on the final evaluation result to improve the recognition rate.



【特許請求の範囲】

【請求項 1】 連続的に入力される音声を、区切り信号の入力時点で区切られる区間を評価する処理で認識する音声認識方法において、

区切り信号の入力をスイッチの操作によって行い、区切り信号の入力時点を、区切り信号の発生時点よりも、予め定めるオフセットだけ先行あるいは遅延して判定することを特徴とする音声認識方法。

【請求項 2】 前記音声の認識の処理では、入力される音声を分析し、分析結果に基づき候補となる単語を選択して、評価値を算出し、評価値の最も高い候補が入力される単語であると判定し、

判定の際に、評価値を、該単語の音声期間に対する前記区切り信号の入力時点の判定結果に従って、変更することを特徴とする請求項 1 記載の音声認識方法。

【請求項 3】 前記区切り信号の入力時点が評価対象となる単語の音声期間の中ほどである場合、該単語の評価値を減少させるように変更することを特徴とする請求項 2 記載の音声認識方法。

【請求項 4】 前記区切り信号の入力時点が評価対象となる単語の音声期間の開始または終了付近である場合、該単語の評価値を増加させるように変更することを特徴とする請求項 2 記載の音声認識方法。

【請求項 5】 前記音声の入力は、記録された音声を再生しながら行うことを特徴とする請求項 1～4 のいずれかに記載の音声認識方法。

【請求項 6】 前記音声の記録時に、区切り信号の記録も行い、前記音声の再生時に、記録された区切り信号を再生しての入力を行うとともに、前記スイッチの操作による区切り信号も入力することを特徴とする請求項 5 記載の音声認識方法。

【請求項 7】 前記区切り信号の入力時点は、前記スイッチに対する ON 操作と OFF 操作とに対応して、それぞれ判定することを特徴とする請求項 1～6 のいずれかに記載の音声認識方法。

【請求項 8】 請求項 1～7 のいずれかに記載の音声認識方法を、コンピュータに実行させるためのプログラムを記録したコンピュータ読取り可能な記録媒体。

【請求項 9】 連続的に入力される音声を、区切り信号の入力時点で区切られる区間を評価する処理で認識する音声認識装置において、

区切り信号の入力操作のためのスイッチと、スイッチの入力操作によって発生する信号波形を整形し、ON/OFF の判定を行う ON/OFF 判定回路と、

ON/OFF 判定回路の判定結果に基づいて、スイッチの入力操作による区切り信号の入力時点を、ON/OFF 判定回路での波形整形による遅延時間をオフセットとして除去し、音声の認識処理を行う音声認識回路とを含

むことを特徴とする音声認識装置。

【請求項 10】 請求項 1～7 のいずれかに記載の音声認識方法で音声認識を行う音声認識手段を備える携帯電話機であって、

該音声認識手段による音声認識処理を能動化させる状態と、停止させる状態とを切替える切替手段と、

切替手段によって、音声認識処理が能動化されているときに、携帯電話機本体の予め定める入力装置への入力操作結果を、区切り信号の入力として処理する入力処理手段とを含むことを特徴とする携帯電話機。

【請求項 11】 ヘッドセットを接続するためのヘッドセット端子を備え、

前記切替手段は、該ヘッドセット端子にヘッドセットが接続されることを検出することによって、前記音声認識手段の音声認識処理を能動化することを特徴とする請求項 10 記載の携帯電話機。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、連続的に入力される音声を、区切りを評価して認識する音声認識方法および装置、ならびにそのような音声認識を行う機能が備えられる携帯電話機に関する。

【0002】

【従来の技術】情報の入力を音声で行うためには、音声認識が必要である。音声認識においては、単語認識を行うために単語全体のパターンを用いる方法、音韻や音節などの単語構成要素の単位を用いる方法、これらを併用して単語認識する方法などが知られている。

【0003】通常の会話などのように、入力音声を自然に連続的に発声させると、単語間で調音結合が起り、単語を単独で発声させた場合の音声パターンから変形してしまう。このため、連続的に発声される単語を認識する場合には、調音結合の影響を加味して作成される辞書を用いると、認識率を向上させることができるとされている。このような音声認識についての先行技術は、たとえば特開平 8-211893 号公報などに開示されている。音声認識結果として、入力された音声に対して最も適切な語を選択するためには、文書としての構造や意味レベルの解析結果などを加えて、最も可能性が高いものを選択する最尤判定などが採用されている。

【0004】しかし、前述の音声認識は、入力された音声のみに基づいて認識を行うために、連続発声した音声から単語を判定するために、種々の辞書を必要とし、音声認識装置の利用者に予め選択した単語を読込ませて学習するなどの準備を行って、認識率向上を図る必要がある。

【0005】また、特開昭 62-166399 号公報には、音声入力以外に補助入力方法、たとえばキーボードのスイッチの押下げなどを使用して、入力音声に対して単語の区切りを入力し、単語が区切れている位置を明確

にして、音声認識過程での単語照合判断の際に利用して、認識を行う方法が開示されている。

【0006】

【発明が解決しようとする課題】単語の区切りを入力しないで音声のみに基づいて音声認識を行う場合は、単語としての区切り位置を判断して、判断された区切り位置を基準にして単語の認識が行われる。しかしながら、単語の区切りの認識自体が困難であり、大量のデータを辞書などとして予め準備しておかなければ認識率を向上させることはできない。

【0007】特開昭62-166399号公報に開示されている先行技術のように、単語の区切りを入力し、入力された区切り信号に従って音声を区切り、区切られた区間毎に単語を認識するようにすれば、単語の区切り自体を認識する必要がなくなり、区切られた区間について認識するために必要な辞書などのデータ量を少なくすることもできる。

【0008】しかしながら、特開昭62-166399号公報に開示されているような区切りの入力方法では、音声の入力と区切りの入力との時間的なずれが考慮されてい
ない。音声入力を行いながら区切り信号を入力する際には、入力者に依存して区切り信号の入力タイミングが異なる。入力者は、単語の区切りを意識して入力操作を行うけれども、単語の発声が終了することを認識した時点で、実際には単語の発声を終了していない段階で区切り信号の入力操作を開始したり、単語の発声が終了してから区切り信号を入力すべきことを認識して区切り信号の入力操作を開始したりする傾向が、個人によって変化する。

【0009】また、区切り操作のためにスイッチを使用すると、機械的な接点がONやOFFする時点で、チャタリングと呼ばれるような振動現象を発生し、ONまたはOFFを確実に1回だけ入力することができず、短時間に複数回のON/OFFが繰返されてしまう。このような繰返し信号は、バウンスノイズと呼ばれ、電氣的な波形整形によって、デバウンス処理が施される。そのようなデバウンス処理は、時間的な遅れを伴ってしまうので、波形整形後のON/OFFの判断時点は、実際の入力操作が行われた時点よりも遅れてしまう。このようにして時間的に遅れた区切りで単語の認識を行うと、正しい認識が行えなくなってしまう。

【0010】また、他人が発声する音声に対して区切りをリアルタイムで同時に入力することは、話の内容の予測がつかないために困難である。また、音声認識において、区切りの位置間で入力信号を切取って認識すると、語頭や語尾が切れて、認識そのものに影響を及ぼしてしまう。また、必ず区切り信号によって区切られた区間を単語として認識しようとする、利用者は単語毎に区切り信号を入力しなければならず、音声の発声でも区切りを意識して発声しなければならなくなって、迅速な音声

入力を行うことができなくなってしまう。さらに、区切りの入力を、音声入力用のマイクロホンと一体になった筐体に取り付けるスイッチによって行う場合には、スイッチを押下げ操作する音もマイクロホンから入力されてしまい、入力音声に対して雑音が混じってしまう。また、スイッチを入力操作する早さは、スイッチの押圧力に依存し、迅速に区切り入力を行うためには、軽い押圧力で入力することができる方が好ましい。

【0011】携帯電話装置などでは、携帯性から全体的な小型化が要求され、スイッチなどの数も少なく、かつ小型で必ずしも操作性はよくない。このために、音声認識機能を備えれば、限られた入力スイッチで多くの機能を利用することが可能となる。しかしながら、携帯電話装置が備えるスイッチは、電話番号などを入力するために用いられ、確実に電話番号などを入力するためにはそれなりの押圧力を必要とするように形成する必要がある。軽い押圧力で入力可能にすると、1つの数字を入力したつもりで、同じ数字を複数回入力してしまうようなことが生じるからである。このため、携帯電話機では、音声入力のための区切り入力用と、携帯電話機本来の入力用のスイッチとを共通使用することが困難である。音声入力用に専用のスイッチを設けることは、携帯電話機ではさらに困難である。

【0012】本発明の目的は、区切り信号の入力を適切に行って、音声認識の精度を向上させることができる音声認識方法および装置ならびに携帯電話機を提供することである。

【0013】

【課題を解決するための手段】本発明は、連続的に入力される音声を、区切り信号の入力時点で区切られる区間を評価する処理で認識する音声認識方法において、区切り信号の入力をスイッチの操作によって行い、区切り信号の入力時点を、区切り信号の発生時点よりも、予め定めるオフセットだけ先行あるいは遅延して判定することとを特徴とする音声認識方法である。

【0014】本発明に従えば、連続的に入力される音声を、スイッチの操作によって入力される区切り信号で区切られる区間を評価する処理で認識する。区切り信号の入力時点は、区切り信号が実際に発生した時点に対して予め定めるオフセットだけ先行あるいは遅延するように変更した時点で判定するので、発声とスイッチ操作との間の操作者によって異なる時間的なずれや、スイッチの信号処理などによる遅れなどの影響を除去して、適切な区切りを行い、音声の認識精度を向上させることができる。

【0015】また本発明の前記音声の認識の処理では、入力される音声を分析し、分析結果に基づき候補となる単語を選択して、評価値を算出し、評価値の最も高い候補が入力される単語であると判定し、判定の際に、評価値を、該単語の音声期間に対する前記区切り信号の入力

時点の判定結果に従って、変更することを特徴とする。

【0016】本発明に従えば、音声認識では、入力される音声の分析結果に基づいて候補となる単語を選択して評価値を算出し、該単語の音声期間に対する区切り信号の入力時点の判定結果で、単語の評価値を変更し、最も高い候補が入力される単語であると判定する。区切り信号を入力しなくてもある程度までは単語の認識を行うことができ、区切り信号を入力すればより高精度で音声認識を行うことができる。

【0017】また本発明は、前記区切り信号の入力時点が評価対象となる単語の音声期間の中ほどである場合、該単語の評価値を減少させるように変更することを特徴とする。

【0018】本発明に従えば、区切り信号の入力時点が評価対象となる単語の発声期間の中ほどである場合にはその単語の評価値を減少させるので、区切りの前後にわたって1つの単語として認識される誤認識が生じる可能性を低減することができる。

【0019】また本発明は、前記区切り信号の入力時点が評価対象となる単語の音声期間の開始または終了付近である場合、該単語の評価値を増加させるように変更することを特徴とする。

【0020】本発明に従えば、区切り信号の入力時点が評価対象となる単語の開始または終了付近である場合にはその単語の評価値を増加させるので、単語の区切りと区切り信号の入力時点とが一致する単語が選択される可能性を高め、認識率を向上させることができる。

【0021】また本発明で前記音声の入力は、記録された音声を再生しながら行うことを特徴とする。

【0022】本発明に従えば、会議などの録音や口述筆記などの録音から音声を再生する間に、区切りを入力して音声認識を行わせることができるので、会議録音や口述筆記などからの文書作成を容易に行うことができる。

【0023】また本発明は、前記音声の記録時に、区切り信号の記録も行い、前記音声の再生時に、記録された区切り信号を再生しての入力を行うとともに、前記スイッチの操作による区切り信号も入力することを特徴とする。

【0024】本発明に従えば、音声の記録時に区切り信号も記録しておき、音声の再生時に区切り信号も再生して音声認識を行いながら、さらに区切り信号を追加して入力することができる。記録された音声を再生しながら区切り信号を追加して入力することによって、音声の誤認識を訂正することができる。

【0025】また本発明で前記区切り信号の入力時点は、前記スイッチに対するON操作とOFF操作とに対応して、それぞれ判定することを特徴とする。

【0026】本発明に従えば、スイッチに対する入力操作では、ON操作とOFF操作とでそれぞれ区切り信号の入力時点と判定するので、スイッチの1回分の入力操

作で2回分の区切り信号を入力させることができ、早口で音声を入力しても、容易に対応する速度で区切り信号を入力することができる。

【0027】さらに本発明は、前述のいずれかに記載の音声認識方法を、コンピュータに実行するためのプログラムを記録したコンピュータ読取り可能な記録媒体である。

【0028】本発明に従えば、記録媒体に記録されているプログラムをコンピュータに読取らせて、音声認識の際に区切り信号の入力時点を適切に判定し、認識精度を向上させることができる。

【0029】さらに本発明は、連続的に入力される音声を、区切り信号の入力時点で区切られる区間を評価する処理で認識する音声認識装置において、区切り信号の入力操作のためのスイッチと、スイッチの入力操作によって発生する信号波形を整形し、ON/OFFの判定を行うON/OFF判定回路と、ON/OFF判定回路の判定結果に基づいて、スイッチの入力操作による区切り信号の入力時点、ON/OFF判定回路での波形整形による遅延時間をオフセットとして除去し、音声の認識処理を行う音声認識回路とを含むことを特徴とする音声認識装置である。

【0030】本発明に従えば、連続的に入力される音声を区切り信号の入力時点で区切られる区間を評価する処理で認識する音声認識装置は、スイッチとON/OFF判定回路と音声認識回路とを含む。区切り信号の入力操作のためのスイッチへの入力操作によって発生する信号波形は、ON/OFF判定回路によって整形され、ON/OFFの判定が行われる。音声認識回路は、ON/OFF判定回路の判定結果に基づいて、スイッチの入力操作による区切り信号の入力時点、予め定めるオフセット時間だけ変更して、音声認識処理を行うので、ON/OFF判定回路での波形整形による時間的な遅れや、発声とスイッチの入力操作との間の時間的なずれなどを補償して、音声に対してスイッチの入力時点を精度よく判定し、入力時点に基づく区間毎の音声認識精度を高めることができる。

【0031】さらに本発明は、前述のいずれかに記載の音声認識方法で、音声認識を行う音声認識手段を備える携帯電話機であって、該音声認識手段による音声認識処理を能動化させる状態と、停止させる状態とを切替える切替手段と、切替手段によって、音声認識処理が能動化されているときに、携帯電話機本体の予め定める入力装置への入力操作結果を、区切り信号の入力として処理する入力処理手段とを含むことを特徴とする携帯電話機である。

【0032】本発明に従えば、携帯電話機には区切り信号の入力時点を適切に判定して音声入力の精度を高めることができる音声認識手段が備えられ、音声認識手段の音声認識処理を能動化させる状態と停止させる状態とを

切換える切換手段を含む。さらに、入力処理手段を含み、入力処理手段は切換手段によって音声認識処理が能動化されているときに、携帯電話機本体の予め定める入力操作結果を区切り信号の入力として処理する。携帯電話機での音声認識が区切り信号を適切に判定して可能となり、認識精度が高まるので、携帯電話機が有する多くの機能を容易に利用することが可能となる。

【0033】また本発明は、ヘッドセットを接続するためのヘッドセット端子を備え、前記切換手段は、該ヘッドセット端子にヘッドセットが接続されることを検出することによって、前記音声認識手段の音声認識処理を能動化することを特徴とする。

【0034】本発明に従えば、ヘッドセット端子にヘッドセットを接続することによって音声認識手段の音声認識処理が能動化され、音声による入力が可能となる。音声の入力にはヘッドセットを利用することができ、区切り信号の入力には携帯電話機本体に設けられている入力装置を利用することができるので、スイッチ操作によって発生する音が入力される音声に対する雑音となって入力されることはなく、雑音による音声認識精度の低下を防ぐことができる。切換手段は、ヘッドセットをヘッドセット端子に接続することによって音声認識手段の音声認識処理を能動化させるように切換えるので、簡単な操作で切換えを行うことができる。

【0035】

【発明の実施の形態】図1は、本発明の実施の一形態としての音声認識装置100の概略的な構成を示す。区切り入力部101は、スイッチへの入力操作に基づく信号を電気的に判定し、スイッチのON/OFF、すなわちスイッチがON状態であるかOFF状態であるかを判定する。スイッチからの信号電圧を、閾値で判定しただけでは、機械的な接点でONやOFF時点でバウンスによるノイズが生じるため、デバウンス処理を等して波形を整形する。このため、スイッチ信号に遅れが生じるので、これを時間的に調整する必要がある。ここでは、40msの遅れが生じる。

【0036】また、スイッチが押釦スイッチなどである場合には、スイッチの入力操作は、人が指で押圧することによって、ON状態となり、指を離すとOFF状態となる。区切りとしての入力時点の判定方法としては、スイッチの1回のONとOFFとの組合せで、区切りが1つ入力されると判定する方法と、スイッチのONに対応して別の区切りを1つ入力すると判定する方法とがある。ONとOFFとに対応してそれぞれ別に区切りを1つずつ入力すれば、同一の操作で2倍の量の区切りを入力することができるので、早口の音声にも追従しやすい。ただし、ONとOFFとのタイミングを単語の区切りに合わせるためには、入力操作に熟練を要する。

【0037】音声入力部102は、マイクロホン、フィ

ルタ、アナログ／デジタル変換器などから成り、マイクロホンで音響信号から電気信号に変換された音声入力を、フィルタによってノイズ成分の除去を行う。フィルタを通過した音声入力は、アナログ／デジタル変換器でデジタルデータ列に変換される。

【0038】分析照合部103は、音声入力部102から与えられるデジタルデータ列から特徴を抽出する。単語パターン記憶部104は、フラッシュROMによって形成され、単語の特徴パターンが記憶される。分析照合部103では、Fast Fourier TransformからFFTと略称される高速フーリエ変換分析や、Linear Predictive Coefficient/CodingからLPCと略称される線形予測分析などの周波数分析が一定周期毎に行われる。なお、Partial auto-CORrelation coefficientからPARCORと略称される偏自己相関分析は、LPCと等価である。このような分析による時系列データが、単語パターン記憶部104に予め記憶されている単語の特徴パターンと比較される。このとき、時間方向は単語の継続時間長によって正規化されて比較される。この比較結果で、単語候補とその評価値としての有意度とが、その存在時刻とともにデータ列として出力される。

【0039】候補判定部105は、分析照合部103から出力される個々の単語候補に対し、区切り入力部101からの信号が、その単語の存在時刻内に存在しているか否かを判断する。単語の存在時刻内に区切り信号が存在する場合は、その単語候補の有意度を変更する。区切り信号が単語の存在する時間的中心に近いほど有意度は引下げられ、開始または終了時刻に近いほど有意度は引上げられる。

【0040】構文・意味解析評価部106は、候補判定部105から与えられる複数の候補を、文法規則、単語辞書、品詞分析などを使用して意味解析し、それぞれの単語候補の有意度とともに重み付け評価を行って、最も評価値が高いものを評価結果として出力する。

【0041】分析照合部103、候補判定部105および構文・意味解析部106は、デジタル信号プロセッサ(DSP)によって、実現され、予め作成されるプログラムに従って処理が行われる。前述のように、区切り入力部101では、デバウンス処理のために40msの時間的な遅れが生じる。また、音声入力の話者とスイッチへの入力操作とを同一人が行うにしても、音声入力とスイッチへの入力操作との間には時間的なずれが生じる。区切り入力部101での遅れや入力操作の時間的なずれを補償するために、オフセット調整部107が設けられる。音声認識装置100では、内部時計108によって与えられる音声入力時点を基準として、区切り入力部101からの区切り信号がオフセット調整部107で時間的に調整され、候補判定部105で、単語候補の存在時刻と区切り信号の入力時点とに基づく有意度の変更が行われる。

【0042】図2は、音声入力波形と区切り入力との関係の一例を示す。横軸の数値は、経過時間を示す。「…音声入力…」という入力が行われる場合、単語の区切り時点、すなわち「音声」と「入力」との単語の境界あるいは開始や終了時点で区切り入力をスイッチを押す操作によって行う。開始時点の区切り入力をA、境界時点の区切り入力をB、終了時点の区切り入力をCとして示す。

【0043】図2の音声波形は、正規化された時間に対して、「O」、「N」、「S」、「E」、「I」、「N Y」、「U」、「R Y」、「O」、および「K」の音素の列として、「音声入力」が分離して解釈される。

【0044】図3に示す単語候補では、候補番号1として、「O」、「N」、「S」、「E」および「I」に対応する波形から、「音声」の単語が選択され、開始時刻が「O」の開始時刻である零と判断され、終了時刻が「I」の終了時刻である310と判断される。各音素の評価を合計して、有意度は95であると判断される。候補番号2は、「O」、「N」、「S」、および「E」の音素の組合せから、「恩師」の単語が候補として選択され、開始時刻は「O」の開始時刻である0となり、終了時刻は「E」の終了時刻200となる。この場合、

「E」の音素が、必ずしも単語と適合していないので、有意度は少し下がり、90となる。候補番号3としては、「I」、「N Y」および「U」の組合せから、「移入」の単語が選択される。「I」の開始時刻270が単語の開始時刻となり、「U」の終了時刻460が単語としての終了時刻となる。単語と音素の組合せとの一致の程度が高いので、有意度は97と高くなる。候補番号4としては、「N Y」、「U」、「R Y」、「O」および「K」の音素の組合せで「入浴」の単語が選択される。開始時刻は「N Y」の開始時刻である320となり、終了時刻は「K」の終了時刻である560となる。「R Y」の部分の一致の程度が小さくなるので、有意度は85となる。候補番号5としては、「N Y」、「U」、「R Y」、「O」および「K」の音素の組合せで、「入力」が選択される。開始時刻および終了時刻は、候補番号4の「入浴」と同様である。「R Y」の部分の一致の程度が高くなるので、有意度は90となる。候補番号6としては、「R Y」、「O」および「K」の組合せで、「録」の単語が選択される。開始時刻は、「R Y」の音素の開始時刻である480となり、終了時刻は「K」の終了時刻である560となる。「R Y」の部分の一致の程度が下がるので、有意度は88となる。

【0045】図3は、図2に示すような音声入力波形に対して、分析照合部103での解析結果としての単語候補を示す。開始時刻および終了時刻は、単語の継続時間長に基づいて正規化し、単語パターン記憶部104に記憶されている単語パターンと比較した音声の存在時間を示す。評点は、単語パターンと入力音声とを比較した結

果得られる評価を示す。

【0046】図4は、音声認識装置100のユーザによる入力操作で入力される区切りを示す。話者と入力操作者とは同一であり、単語の区切りと入力操作とは時間的なずれがないものとする。区切り入力A、B、Cの検出時刻から、前述の処理遅れ時間である40mSを差引いて、修正後時刻を得ている。区切り入力の検出は、ONが入力される時点として行っている。

【0047】図2に示すように、区切り入力Aが入力され、図3の候補番号1の「音声」と候補番号2の「恩師」は、開始時刻が区切り入力Aに一致しているので、これらの単語の有意度は5増加し、それぞれ100および95となる。区切り入力Bの入力時点付近には、候補番号4の「入浴」と候補番号5の「入力」の開始時刻が存在しているので、これらの単語の有意度は5増加し、それぞれ90および95となる。さらに、区切り入力C付近に、候補番号4の「入浴」、候補番号5の「入力」および候補番号6の「録」の終了時刻が存在するので、これらの単語の有意度は5増加し、それぞれ95、100および93となる。

【0048】単語の開始時刻や終了時刻と区切り信号の入力時点との比較では、話者の話す速度の変化や、スイッチへの入力操作時間の変動を考慮して、多少の幅をもたせる。この幅は、たとえば1音素の平均継続時間の1/3とする。すなわち、1つの音素の開始時刻または終了時刻から±1/3の間に入力されれば、「開始時刻付近」または「終了時刻付近」に入力されたものとみなすこととする。図2に示す音声入力波形で、区切り入力Bの入力時点の雑音までの音素を、3音素ずつとって平均的な音素継続時間を算出する。区切り入力Bの前の3音素は、「S」、「E」および「I」であるので、120mSから開始する。区切り入力Bのあとの3音素は、「N Y」、「U」および「R Y」であるので、490mSに終了する。120mSから490mSまでの370mS間には6音素が存在するので、1音素あたりの平均継続時間は62mSとなる。したがって、開始時刻付近と終了時刻付近とは、前後20mS以内となる。

【0049】図5は、本実施形態の音声認識装置100で区切り入力を評価に反映させた音声認識を行う過程を示す。ステップs0から入力開始され、ステップs1で音声入力が音声入力部102に入力され、前述のようにデジタルデータに変換されると、内部時計108からの内部時刻がステップs2で付加される。ステップs3では、分析照合部103での分析照合が行われ、単語候補を表すデータが、その有意度および存在時刻とともに候補判定部105に入力される。候補判定部105には、区切り入力部101からの区切り入力信号が、オフセット調整部107で入力時点の修正をうけて入力される。候補判定部105は、ステップs4で、図3に示すような単語候補を、候補番号順に、各単語候補の存在時

刻中に区切り入力が行われているか否かを判断する。たとえば、候補番号3の単語である「移入」は、270の開始時刻と460の終了時刻との間に存在しているので、区切り入力Bは存在時刻の中ほどに存在している判断される。したがって、ステップs4では、単語候補の存在時刻中に区切り入力があると判断され、ステップs5に移って存在時刻の中ほどに区切りがあると判断される。この結果、ステップs6で、有意度が20減少し、修正後の有意度は77となる。

【0050】ステップs5で、存在時刻の中ほどに区切りが無いと判断されるとき、およびステップs6の有意度の減少が終了したあとは、ステップs7で、存在時刻の開始時付近に区切りが存在しているか否かを判断する。存在しているときには、ステップs8で有意度を5増加させる。ステップs7で存在時刻の開始時付近に区切りが存在していないと判断されるとき、またはステップs8で有意度を増加したあとは、ステップs9で、存在時刻の終了時付近に区切りが存在しているか否かを判断する。存在していると判断されるときには、ステップs10で有意度を5増加させる。ステップs9で存在時刻の終了時付近に区切りが存在していないと判断されるとき、またはステップs10での有意度の増加のあとは、ステップs11で、図3の全候補についての判定が終了しているか否かを判断する。残っている候補があれば、ステップs4に戻り、ステップs4からステップs11までの手順を繰返す。

【0051】ステップs11で、全候補についての判定が終了していると判断されるときには、ステップs12で、候補判定部105の判定結果が構文・意味解析評価部106にわたされ、意味解析が行われる。ステップs13では、ステップs12の意味解析結果と、各単語候補の有意度とが重み付け評価される。この有意度には、ユーザの入力した区切りの有無による再評価が反映されている。ステップs14では、ステップs13での重み付け評価に従って、最も評価値の高い単語候補が認識結果として出力され、ステップs15で1つの音節としての音声認識手順が終了する。

【0052】図6は、本発明の実施の他の形態としての音声認識装置200の概略的な構成を示す。本実施形態の音声認識装置200で、図1に示す音声認識装置100に対応する部分には同一の参照符を付し、重複する説明を省略する。本実施形態では、音声入力部102に入力される音声入力は、一旦記録／再生部201に記録される。記録／再生部201は、音声入力部102でデジタル化され、内部時計108によって与えられる内部時刻が付加された音声データを、記憶素子に記憶する機能を有する。記憶素子は、たとえばフラッシュROMやRAMなどの半導体メモリ、磁気テープや磁気ディスク、MD、CD-RWなどの記録媒体を用いることができる。記録／再生部201は、記憶素子に記憶された音声

データを再生して出力する機能も有する。音声出力部202は、デジタル化された音声データを、デジタル／アナログ変換器でアナログ信号に変換し、フィルタでノイズを除去し、アンプで増幅して、スピーカで電気信号から音響信号に変換して、音響信号として出力する。

【0053】図1の音声認識装置100では、音声の入力と区切りの入力とを同時に行うけれども、本実施形態の音声認識装置200では、一旦録音したあとで、音声出力部202からの音声出力を聞きながら、区切り入力部101から区切りを入力することができる。これによって、会議の議事録を作成するような場合に、会議の録音を聞きながら区切りを入力して、音声認識の認識率を向上させることができる。会議中に区切り入力を行うことは、他の人の発言であり、自身の発言でない場合には非常に難しい。単語が終了していることは、次の単語に入った時点で明確になり、その時点で区切り信号を入力すると前の単語の終了時刻からは遅れ、次の単語の中間で入力されたこととなって次の単語の認識を誤らせてしまうおそれがある。

【0054】なお、本実施形態の音声認識装置200では、音声入力部102でデジタル化された音声データを記録／再生部201で記録するようにしているけれども、磁気テープやMDなどに録音した音声も、図1の実施形態の音声認識装置100の音声入力部102に入力しながら、区切り入力部101に区切りを入力するようにして、会議の議事録の作成や口述筆記などを行うこともできる。また、音声認識処理装置100の音声入力部102に、放送や電話、あるいはライン入力などで与えられる電気的な音声信号を与えて、それを再生しながら区切りを入力し、音声認識を行わせることもできる。

【0055】図7は、本発明の実施のさらに他の形態としての音声認識装置300の概略的な構成を示す。本実施形態の音声認識装置300で、図1の音声認識装置100や図6の音声認識装置200に対応する部分には同一の参照符を付し、重複する説明を省略する。図7に示す音声認識装置300では、①区切り入力も音声と同時に記録、②再生時にさらに区切り入力を追加可能、③句読点を入力するスイッチを設けている。記録／再生部301では、図6に示す音声認識装置200の記録／再生部201に対して、区切り入力に使用する区切り入力スイッチの数だけ記憶素子の記憶ビット数を増やし、音声と同時に区切り入力を記録することができるように変更してある。区切り追加部302は、記録／再生部301の再生時に出力する記憶されている区切りと、新たにスイッチから入力される区切り入力とを加えて、両方を候補判定部105に提供する。句読点入力部303は、区切り入力部101と同様に構成され、スイッチの入力操作に従って句読点の入力を行う。句読点入力部303からの出力は、構文・意味解析評価部306に与えられる。また、句読点は、単語の区切りとしての意味も有す

るので、区切り追加部302を介して候補判定部105にも与えられる。本実施形態のオフセット調整部307は、音声信号に対するFIFOメモリを備え、スイッチからの信号に基づいて、スイッチ入力や音声入力に遅延を与え、それらの時間的ずれを補償する。

【0056】本実施形態の構成によれば、①区切り入力も音声と同時に記録／再生部301に記録し、区切り追加部302によって②再生時にさらに区切り入力を追加して図4に示すような表を作成することができる。③句読点を入力するスイッチを設け、句読点入力部303からの出力を候補判定部105に与えて、区切り入力と同様に単語区切りとして使用するとともに、構文・意味解析評価部306に与えて、文章がどこで終わるかを明示して連続発声した音声でも構文解析を区切って実行することができる。

【0057】区切り入力に生じる時間的遅延を補償するオフセット調整部307では、音声信号はその波形データのままFIFOで遅延させ、分析照合部103に入力する。区切り入力部101からの出力は、図4に示すような表で与えられ、オフセット調整部107では、この表の修正の時刻を加減演算して書き直す処理として実現される。これによって、話者に依存して起こる発声音声とスイッチ入力とのタイミングのずれを補正することができる。また、会議議事録の作成のように、話者とスイッチ入力操作者が異なる場合には、さらに遅延が増加するので、これを調整することもできる。

【0058】図8は、図7の実施形態の音声認識装置300を用いて行う音声認識の概要を示す。図8(a)は、音声の記録手順を示す。ステップa0から記録を開始し、ステップa1では音声が入力される。ステップa2では、音声入力部102からの出力が記録／再生部301に記録される。ステップa3では、区切り入力部101からのスイッチ入力の有無が判断される。スイッチ入力があれば、ステップa4で、区切り入力が記録／再生部301に記録される。ステップa3でスイッチ入力がないと判断されるとき、またはステップa4の記録が終了すると、ステップa5で音声入力が終了しているか否かを判断する。終了していなければ、ステップa1に戻り、ステップa1からステップa5までの手順を繰り返す。ステップa5で音声入力が終了していると判断されれば、ステップa6で記録の手順を終了する。

【0059】図8(b)は、記録／再生部301に記録されている音声を認識する手順を示す。ステップb0で再生を開始し、ステップb1では区切りが再生されているか否かを判断する。区切りが再生されていないときには、ステップb2で区切り入力部101に対してスイッチ入力操作による区切り入力が行われているか否かを判断する。ステップb1で区切りが再生されていると判断されるとき、またはステップb2で区切りが入力されていると判断されるときには、ステップb3で、オフセッ

トの調整が行われる。ステップb3でのオフセットの調整後、またはステップb2で区切り入力も入力されていないと判断されるとき、ステップb5、特にそのステップs3からステップs15までと同様な、音声認識処理が行われる。ステップb5では、記録された音声が入力されているか否かを判断する。残りがあるときには、ステップb1に戻り、ステップb1からステップb5までの手順を繰り返す。ステップb5で、記録の再生が終了していると判断されるときには、ステップb6で音声認識の手順を終了する。

【0060】図9は、本発明の実施のさらに他の形態として、本発明の音声認識の機能を備える携帯電話機600の概略的な構成を示す。携帯電話機600には、無線／通信部601、マイコン602、入力スイッチ603、ヘッドセットジャック604、音声認識手段605、本体マイク606および本体スピーカ607などが備えられている。無線／通信部601は、無線電波の送受信で携帯電話用の公衆回線と通信を行い、無線で送受信するデータと音声との間の変換を行うチャンネルコーデックなどを含む。マイコン602は、電話のプロトコルやユーザインターフェース処理を行う電話処理手段608と、音声認識手段605への音声入力に従って電話処理手段608を制御する入力結果処理手段609を含む機能を実現する。

【0061】音声認識手段605は、図1、図6あるいは図7に示すいずれかの実施形態と同様に音声認識処理を行う。音声認識手段605は、アナログ／デジタル変換器やデジタル信号プロセッサなどが含まれ、音声認識はデジタル信号プロセッサのソフトウェア処理で行われる。デジタル信号プロセッサは、電話通信に必要な音声圧縮と音声伸長機能も実現する。

【0062】ヘッドセットジャック604にヘッドセット610のプラグを差込むと、挿入検出手段611によって検出され、マイコン602に伝えられる。マイコン602は、切換スイッチ612としても機能し、携帯電話機600が備える入力スイッチ603を、音声認識手段605の区切り入力スイッチとして動作するように切換える。音声認識手段605は、図7の実施形態と同様に構成される場合には、入力スイッチ603のうちで句読点入力用のスイッチを音声認識手段605側に切換えるようにすればよい。

【0063】携帯電話機600では、本体の筐体の表面に入力スイッチ603や本体マイク606および本体スピーカ607などが装着されている。区切り入力のために入力スイッチ603を使用すると、本体マイク606から音声入力を行う場合には、キーの入力操作後も拾ってしまう可能性がある。本実施形態では、ヘッドセット610を使用するときに本体マイクではなくヘッドセット610のマイクを使用するので、入力スイッチ603の入力操作音が音声入力に混入することとはなく、また入

力スイッチ 603 の操作も簡単になる。また、表示部にタッチパネルを設け、これを区切り入力に使用すれば、軽いタッチで入力も可能となる。

【0064】携帯電話機 600 では、携帯性が重視されるので、入力スイッチ 603 としてあまり多くのキースイッチを設けることはできない。音声認識手段 605 を備えていれば、音声による制御や音声による入力が可能となり、電子メールの文字入力も容易となり、多くの機能を実現することができる。

【0065】本発明の音声認識処理は、図 9 に示すような携帯電話機 600 ばかりではなく、パーソナルコンピュータ、ワードプロセッサあるいは PDA などの情報機器や、各種電気製品の制御や入力に使用することができる。入力や制御を、リモコンで行う場合にも同様に適用することができる。また、パーソナルコンピュータや PDA などに、本発明を適用した音声認識処理のためのプログラムを記録媒体やネットワークからのダウンロードで読込ませて、音声認識を行わせることもできる。

【0066】

【発明の効果】以上のように本発明によれば、音声認識の精度を向上させるための区切り信号の入力に対し、オフセットを設けて入力時点の判定を行うので、話者や操作者に依存する発声と区切り信号の入力操作との時間的なずれを補正したり、スイッチからの入力信号に対する雑音除去処理などで生じる遅延を補償して、認識精度を向上させ、誤認識を減少させることができる。

【0067】また本発明によれば、区切り信号の入力によって、音声認識の候補となる単語の評価値を変更し、認識精度を高めることができる。

【0068】また本発明によれば、区切り信号の入力によって、区切り信号の入力時点の前後にわたるような単語で認識される可能性を減少させ、誤認識を避けることができる。

【0069】また本発明によれば、区切り信号の入力によって、単語の区切りと一致する候補への認識の可能性を高め、認識率を向上させることができる。

【0070】また本発明によれば、一旦記録された音声に対する精度の高い認識が可能となる。たとえば会議などで録音したあと、区切りを独立に後から入力することができるので、話者と区切り信号の入力操作者とが異なってもよく、誰にでも音声認識のための区切り信号の入力操作を適用することができる。これによって、口述筆記や会議録音から文書作成を行うことが容易になる。

【0071】また本発明によれば、音声を記録しながら区切り信号も記録することができ、さらに音声を再生する際に区切り信号を追加して入力することもできるので、音声に対し区切り信号の入力を一層適切に行うことが可能となり、誤認識の訂正が容易となる。

【0072】また本発明によれば、スイッチへの ON 操作と OFF 操作とに対応して区切り信号の入力時点を判

定するので、スイッチへの 1 回の ON 操作と OFF 操作との組合せで、2 回分の区切り信号の入力を行うことができ、早口で入力される音声信号に対しても、容易に対応する速度で区切り信号の入力を行うことができる。

【0073】さらに本発明によれば、コンピュータを用いての音声認識で、区切り信号の入力時点を適切に判定して、音声認識の精度を向上させることができる。

【0074】さらに本発明によれば、音声信号を入力しながら区切り信号入力のために操作するスイッチからの信号を、波形整形して ON/OFF の判定を行い、区切り信号の入力時点、予め定めるオフセット調整で波形整形による遅れを補償して行うので、音声認識の精度を向上させることができる。

【0075】さらに本発明によれば、携帯電話機で、音声認識処理を、区切り信号の入力時点を適切に判定して行うことができるので、音声認識の認識精度を高め、携帯電話機の機能を音声によって活用することができる。

【0076】また本発明によれば、ヘッドセットを接続して、ヘッドセットから音声入力を行い、携帯電話機本体に備えられる入力装置で区切り信号の入力操作を行うので、入力装置の入力操作に伴う雑音が音声入力に混じることがなく、雑音による音声認識の精度低下を防ぐことができる。

【図面の簡単な説明】

【図 1】本発明の実施の一形態の音声認識装置 100 の概略的な構成を示すブロック図である。

【図 2】図 1 の音声認識装置 100 に入力される音声入力波形と区切り入力のタイミングとの例を示すグラフである。

【図 3】図 1 の分析照合部 103 によって抽出される単語候補の例を示す図表である。

【図 4】図 1 の区切り入力部 101 に入力される区切り入力の検出時刻と、オフセット調整部 107 によって修正された修正後時刻とを示す図表である。

【図 5】図 1 の音声認識装置 100 による音声認識の手順を示すフローチャートである。

【図 6】本発明の実施の他の形態の音声認識装置 200 の概略的な構成を示すブロック図である。

【図 7】本発明の実施のさらに他の形態としての音声認識装置 300 の概略的な構成を示すブロック図である。

【図 8】図 7 の音声認識装置 300 が音声入力と区切り入力を記録する手順を示すフローチャート、および記録された音声および区切り入力を再生しながら、さらに区切り入力を追加して音声認識処理を行う手順を示すフローチャートである。

【図 9】本発明の音声認識処理が実施される携帯電話機 600 の概略的な構成を示すブロック図である。

【符号の説明】

100、200、300 音声認識装置

101 区切り入力部

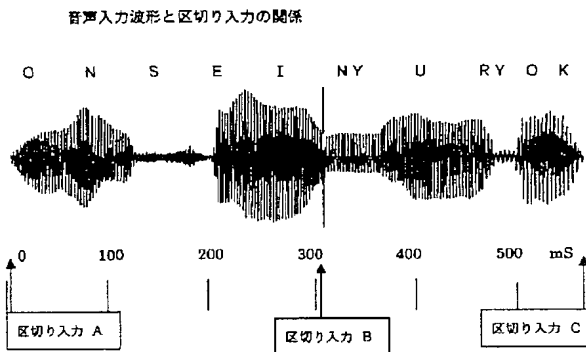
102 音声入力部
 103 分析照合部
 104 単語パターン記憶部
 105 候補判定部
 106, 306 構文・意味解析評価部
 107, 307 オフセット調整部
 108 内部時計
 201, 301 記録／再生部
 202 音声出力部

*

* 302 区切り追加部
 303 句読点入力部
 600 携帯電話機
 603 入力スイッチ
 605 音声認識手段
 606 本体マイク
 610 ヘッドセット
 612 切換スイッチ

【図2】

【図3】



単語候補

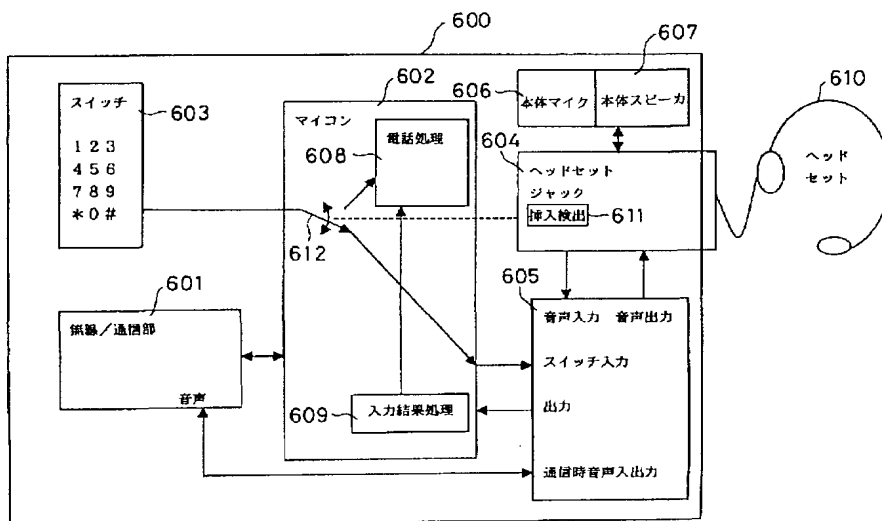
候補番号	単語	開始時刻	終了時刻	評点
1	音声	0	310	95
2	愚痴	0	200	90
3	移入	270	460	97
4	入浴	320	560	85
5	入力	320	560	90
6	録	480	560	88

【図4】

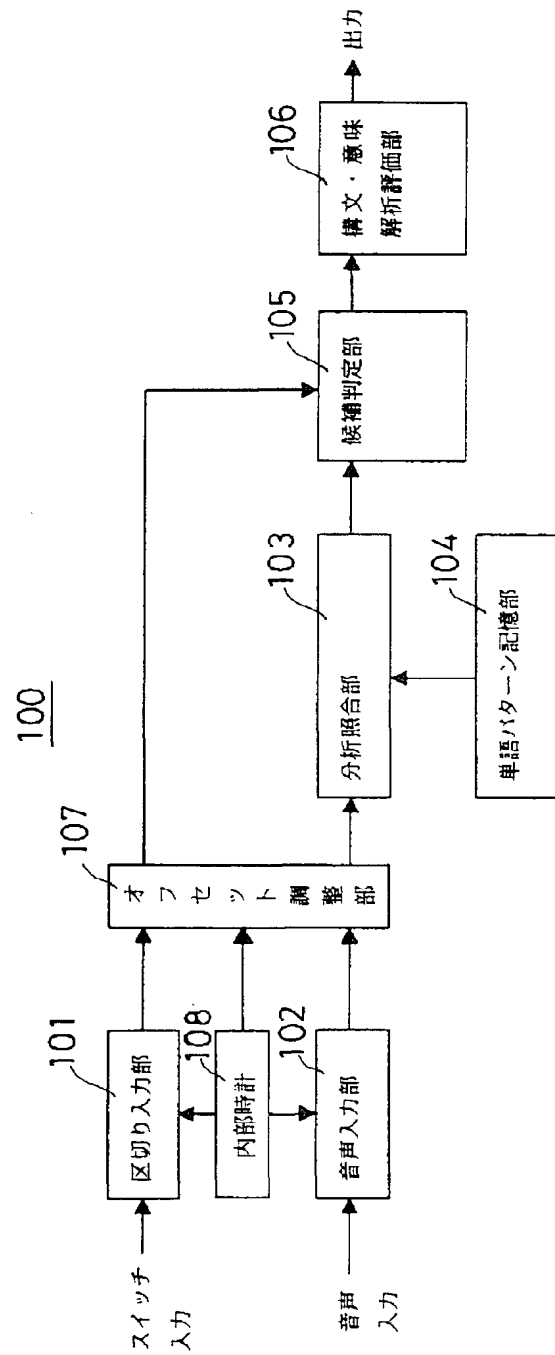
区切り入力

区切り	修正後時刻	検出時刻	検出
A	0	40	ON
B	305	345	ON
C	560	600	ON

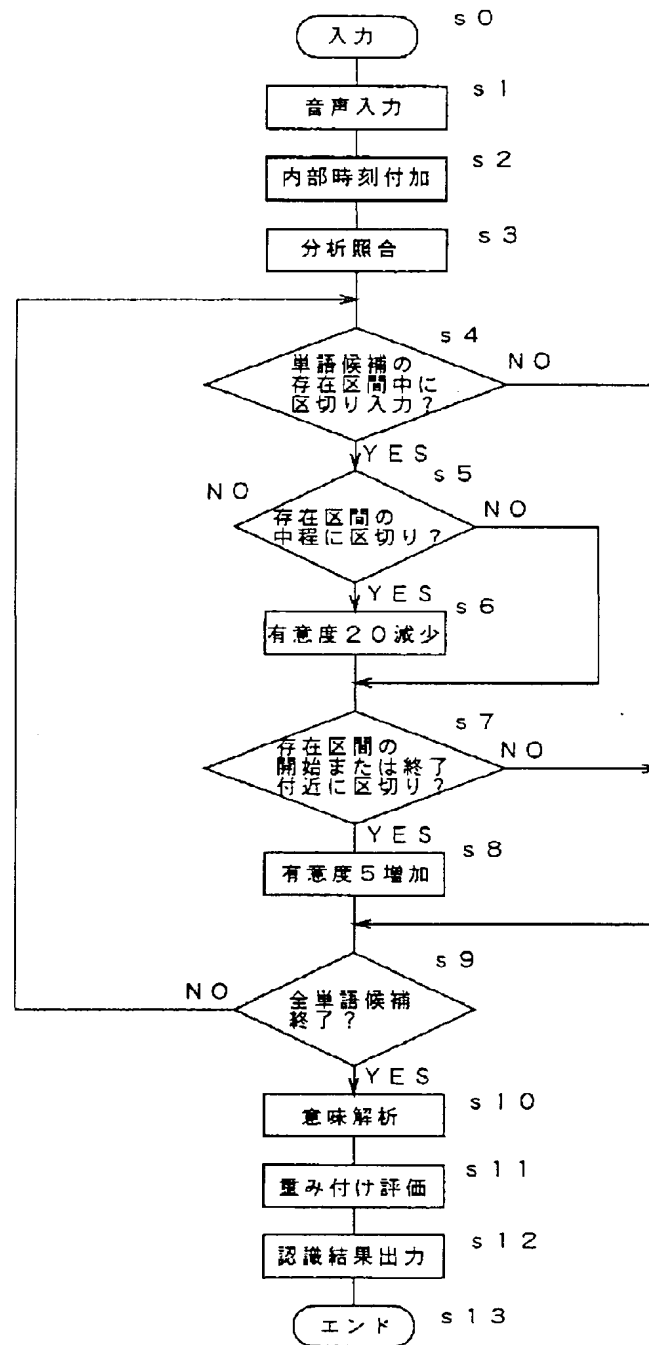
【図9】



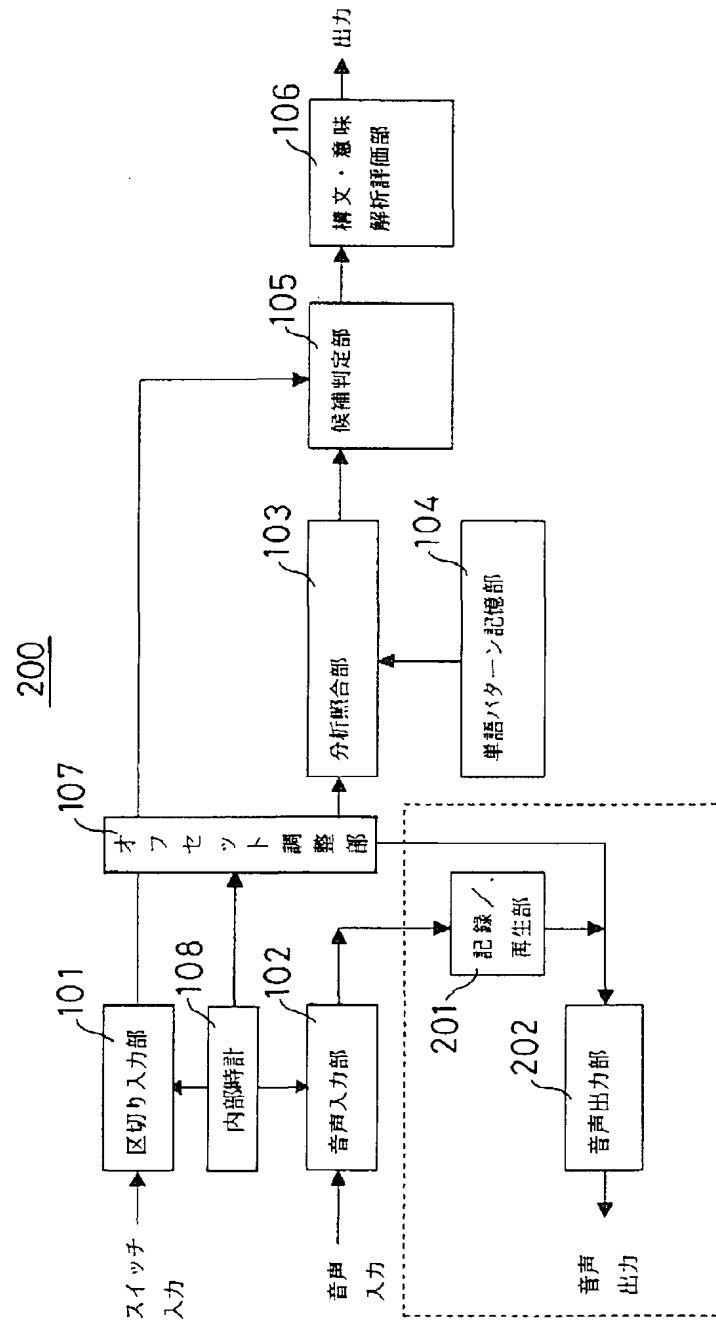
【図1】



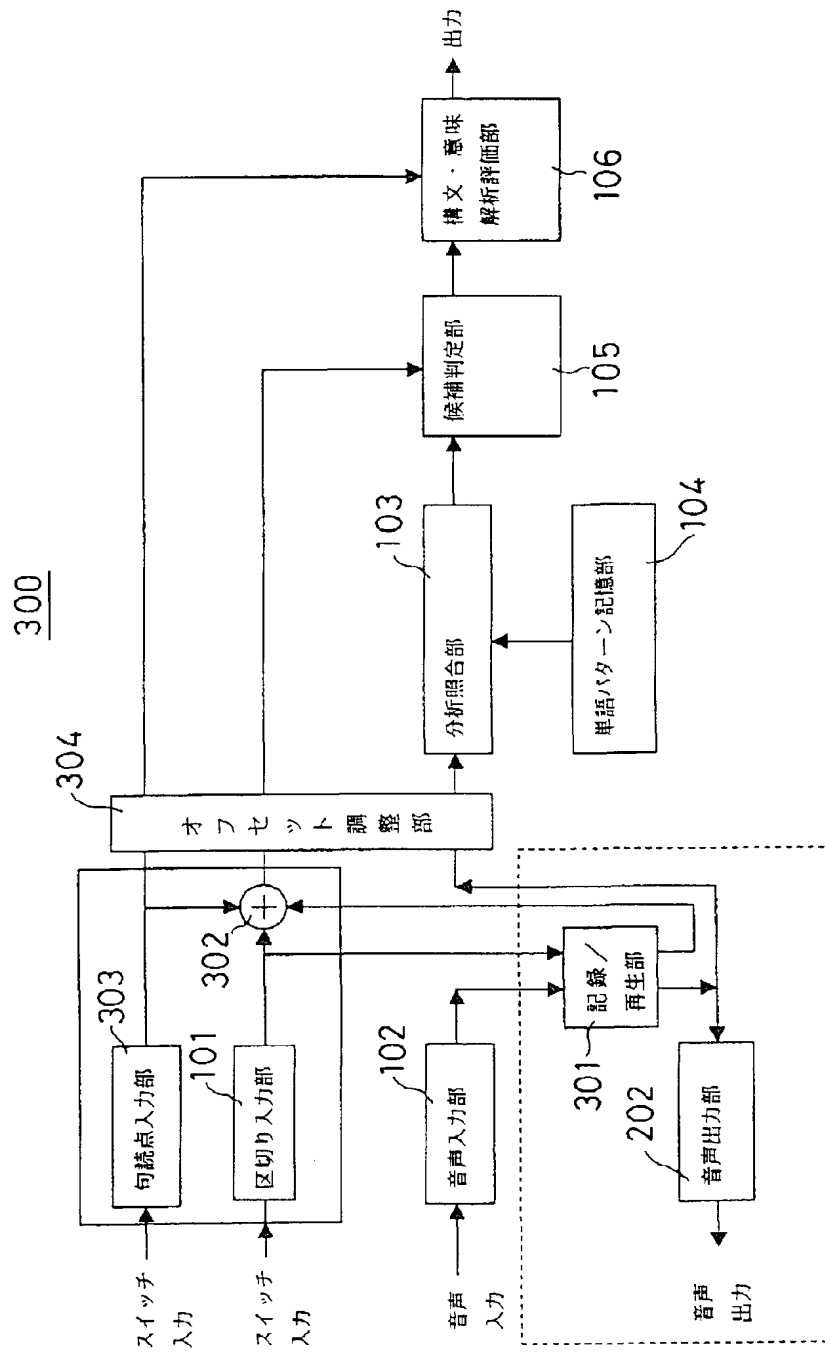
【図5】



【図6】



【図7】



【図8】

